Preparing to Run on Pleiades Broadwell Nodes

To help you prepare for running jobs on Pleiades Broadwell compute nodes, this short review includes the general node configuration, tips on compiling your code, and PBS script examples.

Overview of Pleiades Broadwell Nodes

Each Pleiades Broadwell rack contains 72 nodes; each node contains two 14-core E5-2680v4 (2.4 GHz) processors and 128 GB of memory, providing approximately 4.6 GB per coreâ   slightly less than the ~5.3 GB/core provided by the Haswell nodes.

The Broadwell nodes are connected to the Pleiades InfiniBand network (ib0 and ib1) via four-lane Fourteen Data Rate (4X FDR) devices and switches for internode communication.

The home and Lustre /nobackup filesystems are accessible from the Broadwell nodes.

Compiling Your Code For Broadwell Nodes

Like the Haswell processors, the Broadwell processors support the Advanced Vector Extensions 2 (AVX2) instructions, in addition to AVX (introduced with Sandy Bridge processors), SSE4.2 (introduced with Nehalem processors), and earlier generations of SSE.

AVX2 supports floating point fused multiply-add, integer vector instructions extended to 256-bit, and vectorization gather support, among other features. Your application may be able to take advantage of AVX2, however, not all applications can make effective use of this set of instructions. We recommend that you use the latest Intel compiler, by using the command **`module load comp-intel/2020.4.304,`** and experiment with the following sets of compiler options before making production runs on the Broadwell nodes:

- **`-O2 -xCORE-AVX2`**
- **`-O3-xCORE-AVX2`**

These compiler options generate an executable that is optimized for running on Broadwell and Haswell, but cannot run on any of the older processor types. If you prefer to generate a single executable that will run on any of the existing Pleiades processor types (Sandy Bridge, Ivy Bridge, Haswell, Broadwell, Skylake, and Cascade Lake), try using one of the following sets of compiler options:

- **`-O2 (or -O3)`**
- **`-O2 (or -O3) -axCORE-AVX512,CORE-AVX2 -xAVX`**

The use of **`-axCORE-AVX512,CORE-AVX2`** allows the compiler to generate multiple code paths with suitable optimization to be determined at run time.

If your application does not show performance improvements with **`-xCORE-AVX2`** or **`-axCORE-AVX512,CORE-AVX2 -xAVX (`**as compared with just **`-o2`** or **`-o3`**) when running on Broadwell and Haswell nodes, it is more advantageous to use the executable built with just **`-o2`** or **`-o3`** for production runs on all Pleiades processor types.

TIP: You can add the compiler options **`-ip`** or **`-ipo`** to instruct the compiler to look for ways to better optimize and/or vectorize your code. Also, to generate a report on how well your code is vectorized, add the compiler flag **`-vec-report2`**.
Notes:

- If you have an MPI code, we strongly recommend that you load the **mpi-hpe/mpt** module, which always points to the NAS recommended MPT version.
- Ensure your jobs run correctly on Broadwell nodes before you start production work.

Running PBS Jobs on Broadwell Nodes

A PBS job running with a fixed number of processes or threads should use fewer Broadwell nodes than other types of Pleiades nodes, for two reasons: Broadwell nodes have more cores and more memory.

To request Broadwell nodes, use **model=bro** in your PBS script. For example:

```
 #PBS -l select=xx:ncpus=yy:model=bro
```

Note: If your job requests only **model=bro** nodes, then by default PBS will run your job on either Pleiades or Electra Broadwell nodes, whichever becomes available first. If you specifically want your job to only run on Pleiades, then add **-l site=static_broadwell** to your job request. For example:

```
#PBS -l select=10:ncpus=28:mpiprocs=28:model=bro
#PBS -l site=static_broadwell
```

Similarly, a request to use Electra Broadwell nodes using **model=bro_ele** will by default run on either Pleiades or Electra Broadwell nodes. For more information, see Preparing to Run on Electra Broadwell Nodes.

Cores per Node

There are 28 cores per Broadwell node compared to 24 cores per Haswell, 20 cores per Ivy Bridge, and 16 cores per Sandy Bridge.

For example, if you have previously run a 240-process job with 10 Haswell nodes, 12 Ivy Bridge nodes, or 15 Sandy Bridge nodes, you should request 9 Broadwell nodes (where the first node can run 16 processes while the remaining 8 nodes can run 28 processes each).

```
For Broadwell
#PBS -lselect=1:ncpus=16:mpiprocs=16:model=bro+8:ncpus=28:mpiprocs=28:model=bro

For Haswell
#PBS -lselect=10:ncpus=24:mpiprocs=24:model=has

For Ivy Bridge
#PBS -lselect=12:ncpus=20:mpiprocs=20:model=ivy

For Sandy Bridge
#PBS -lselect=15:ncpus=16:mpiprocs=16:model=san
```

Memory

Except for the Haswell node type, the Broadwell type provides more memory compared to the other Pleiades processor types both on a per node or per core basis.

For example, to run a job that needs 4 GB of memory per process, you can fit 7 processes on a 16-core Sandy Bridge node with ~30 GB/node, 15 processes on a 20-core Ivy Bridge node with ~60 GB/node, 24 processes on a 24-core Haswell node with ~122 GB/node, and 28 processes on a 28-core Broadwell node with ~122 GB/node.

Note: For all processor types, a small amount of memory per node is reserved for system usage. Therefore, the amount of memory available to a PBS job is slightly less than the total physical memory.

Sample PBS Script For Broadwell Nodes

```
#PBS -lselect=10:ncpus=28:mpiprocs=28:model=bro
#PBS -q devel
module load comp-intel/2020.4.304 mpi-hpe/mpt
cd $PBS_O_WORKDIR
mpiexec -np 280 ./a.out
```

For more information about Broadwell nodes, see:

- Pleiades Configuration Details
- Broadwell Processors

Article ID: 530
Last updated: 20 May, 2021
Revision: 44
Systems Reference -> Pleiades -> Preparing to Run on Pleiades Broadwell Nodes
https://www.nas.nasa.gov/hecc/support/kb/entry/530/